Population coding for a reward-modulated Hebbian learning of vergence control

Agostino Gibaldi, Andrea Canessa, Manuela Chessa, Fabio Solari and Silvio P. Sabatini

Abstract—We show how a cortical model of early disparity detectors is able to autonomously learn effective control signals in order to drive the vergence eye movements of a binocular active vision system.

The proposed approach employs early binocular mechanisms of vision and basic learning processes such as synaptic plasticity and reward modulation. The computational substrate consists of a population of modeled V1 complex cells, that provides a distributed representation of binocular disparity information. The population response also provides a global signal to describe the state of the system and thus its deviation from the desired vergence position.

The proposed network, by taking into account the modification of its internal state as a consequence of the action performed, evolves following a differential Hebbian rule. Furthermore, the weights update is driven by an intrinsic signal derived by the overall activity of the population. Exploiting this signal implies a maximization of the population activity itself, thus providing an highly effective reward for the developing of a stable and accurate vergence behaviour.

The efficacy of the proposed intrinsic reward signal is comparatively assessed against the ground-truth signal (the actual disparity) providing equivalent results, and thus validating the approach.

Experimental tests in a simulated environment demonstrate that the proposed network is able to cope with vergent geometry and thus to learn effective vergence movements for static and moving visual targets in realistic situations.

I. INTRODUCTION

The primary visual cortex (V1), in humans and monkeys, is considered the processing substrate for the retinal binocular disparity, and enables both stereopsis and vergence eye movements [17], [4]. While stereopsis is the process that allows for the perception of depth by disparity information, vergence movements are responsible for ensuring the singleness of vision and for maintaining stable fixations. These two mechanisms mutually influence and develop together in a process in which a finer vergence movement helps improving the stereacuity, and a finer stereacuity provides effective information for finer vergence movements.

When an infant starts looking around, both the mechanisms are not yet present in the cortex, and have to be learned. Since the infant has no explicit teacher to guide him, its only “supervision” is gathered from interaction, i.e. from a direct sensorimotor connection with the environment [27]. A plausible learning process should count on a reward given when the eyes fixate an object in the proper manner, thus when the singleness of vision is guaranteed. This process ensures that an infant will learn the correct vergence behaviour [1], [29], and that an adult can adapt the control to the stimulus characteristics [33]. In fact, adaptation in the vergence system is “essential for an organism to maintain optimal visuomotor function” [15].

The retinal binocular disparity is used by the brain as a source of information to gain depth perception and to control the eye movements, in order to actively get a better perception of the scene, on the basis of the characteristics of the scene itself. Indeed, a visual stimulus containing a disparity, like a random dot stereogram (RDS), is equally effective in providing depth perception and in triggering the correct vergence eye movements [17], [4]. From a computational point of view, although the complex cells of the primary visual cortex (V1) are the processing substrate for both stereopsis [18] and vergence [17], these two tasks are carried out by two separate cortical mechanisms. While the former is capable of producing a single perception from two different retinal images, only within a small range of disparities (Panum’s fusional area [32]), the latter allows us to extract a vergence control for large disparities as well. Accordingly, disparity-vergence responses might follow a fast reactive stream that directly involves V1 cells without resorting to a high level interpretation of depth. In this way, the system is brought back to the fusible range in order to ensure again the singleness of vision.

Even though the stereo and vergence mechanisms are supposed to develop in parallel and to refine and calibrate each other, the approaches that jointly address this issues are very seldom [26], [34]. Indeed, early works showed on the one hand how it is in principle possible to learn the receptive fields of simple and complex cells from the statistical properties of the natural images [16], [20], [12] and on the other hand how a network of disparity detectors is an effective substrate to guide vergence eye movements [28], [21]. At an intermediate level, we can start from a large population of already developed V1-like cells and make the network to develop effective vergence control. In [2], a convolutional neural network is trained by exploiting the vergence error (difference between the actual and the desired vergence) as the learning signal. Notwithstanding the effectiveness of the approach, it requires a ground-truth knowledge of the robotic system in relation with the environment, that is unlikely to be available in real world.

In [9], [6], the authors proposed a read-out mechanism
of the response of a population network of V1 complex cells so to specialize it for an effective vergence behaviour. The vergence control is computed through a weighted linear summation of the population response. The weights are obtained by a least mean square (LS) algorithm. The population tuning curves are used as basis function and to approximate a desired behaviour. On the one hand, imposing a behaviour allows the architecture to take full advantage of the resources, so to cope with a wide range of disparities. On the other hand, a behaviour that is externally imposed, requires a complete knowledge of the available resources.

From the perspective of an active vision systems, a key point is that the control should be learned actively through a direct interaction with the environment, not by imposing an external predefined behaviour. In [5] the authors focus on how a learned sensory representation is able to guide vergence movements in a behaving organism, using a biological and unsupervised reward. To this end, the response of a population of complex cells tuned to zero disparity is used to earn the emergence of disparity tuning in a three-layer neural network, obtaining in such a way, a plausible substrate for guiding vergence movements. The stimuli used for the training are one-dimensional (i.e. single image rows), thus simplifying the dimensionality of the problem to horizontal disparity only. The network is able to specialize different servos with high sensitivity for a small disparity and broad curves for large disparities, thereby validating the approach.

The more recent models of [30], [31] propose to use the output of different populations of complex cells, both for vergence control and for the reward. In particular, the reward is computed from the responses of the different populations at each step, thus with no previous knowledge about the environment. However, since they use receptive fields with a vertical orientation ($\theta = 0$), their populations are tuned to the horizontal disparity component only, thus limiting the complexity of the problem. Their approach was extended in [23], where the role of neurons tuned to different orientation has been explored.

It is worth noting that these models rely on a selected set of resources, characterized by specific binocular properties, or specific classes of disparity tuned cells. Much more appealing, in the context of autonomous learning, is to assess whether a network of disparity detectors is able to learn the proper behaviour without any external supervision, and what can be an ideal signal to evaluate the performance of the system and to drive the learning phase. In [7] the authors demonstrated how the activity of a population of disparity detectors, designed following biological inspired specification, can be exploited not only to compute the control for the vergence movement, but also to define the state of the vision system with respect to the environment, and consequently to evaluate the effect (positive or negative) of the movement on the system status. Nevertheless, the proposed architecture is trained by using a particle swarm optimization, that is not likely to be a possible neural strategy.

In this paper we demonstrate how the properties of the neural architecture of disparity detectors implicitly allows for a learning strategy of the control of vergence eye movements, grounding on basic cortical mechanisms. The population of disparity detectors includes a normalization stage that ensures that the population activity is stable and sensitive to retinal disparity only. The vergence control is provided by a linear network whose weights are updated following a Hebbian rule. Since the mechanism has to evaluate the effect of an action, the basic rule has been modified to take into account a temporal asymmetry, and it is driven by an internal reward signal, provided by the overall population activity. Indeed, a correct reward for the actions taken by the system is internally obtained by the overall increase of the population activity, which can be used to intrinsically train the system toward an effective solution.

The paper is organized as follows: Sec. II describes the cortical architecture of V1 complex cells and the role of the divisive normalization, Sec. III describes the implementation of the learning algorithm, based on the population response, Sec. IV analyzes the results obtained both with synthetic and more complex stimuli, and finally in Sec. V we present our conclusions.

II. EFFICIENT CODING OF BINOCULAR DISPARITY

Taking inspiration from neurophysiological evidences [3], our architecture of disparity detector consists of a population of V1-like binocular energy complex cells obtained from the sum of squared responses of quadrature pairs of simple cells. These complex cells, each with its own sensitivity to a particular vectorial disparity $\delta(x) = (\delta_H, \delta_V)$, allow the population to implicitly code, the information of $\delta(x)$. The population is sensible to disparities bounded in a circular subregion of the 2D disparity domain centered around zero. The region radius is directly related to the size of the simple cell’s receptive field (RF), modeled by a complex-valued Gabor filter $h(x; \theta, \psi) = \rho e^{-\frac{1}{2\sigma^2}x^2} e^{i\psi + i2\pi k_0 x}$ where $x$ is the center of the filter on the image plane, $x_0 = [x_0, y_0]$ is the coordinate system rotated by an angle $\theta$ about its center, $k_0$ is the radial peak frequency of the filter, $\rho$ is a normalization constant, and $\psi$ is the phase value. The response of the left and right RFs centered in $x$ in the images $I_{L/R}$ is:

$$r_{L/R}(x; \theta, \psi_{L/R}) = I_{L/R} \ast h_{L/R}(x; \theta, \psi_{L/R})$$

and the response of a modeled binocular complex cell is:

$$r_c(x; \theta, \Delta \psi) = |r_L(x; \theta, \psi_L) + r_R(x; \theta, \psi_R)|^2$$

Scaling up to the 2D case the 1D approach proposed in [4], it is possible to analytically derive the complex cell’s response using a synthetic ideal stimulus, a white Gaussian noise characterized by a constant Fourier power spectrum.
Fig. 1. The 2D response profile of the V1 modeled complex cells to a RDS stimuli (top row), and the approximation obtained by the analytical model (bottom row). The stimulus disparity varies in the range $[-3\Delta, 3\Delta]$ for both $\delta_T$ and $\delta_R$. The parameters used for the two represented cells are: (A) $\theta = 0$ and $\psi = 0$, and (B) $\theta = \pi/4$, $\psi = 3/8\pi$. The bottom insets show the horizontal cross section of the actual responses (black line) and of the analytical model (blue line).

Expanding Eq. 2, the tuning curve approximates to:

$$r_c(x; k_\theta, \Delta \psi) \approx \frac{16\pi^4|\tilde{I}|^2}{\sigma^4} \left[ 1 + \exp \left( -\frac{|\delta(x)|^2}{\sigma^2} \right) \right] + 2\exp \left( -\frac{|\delta(x)|^2}{2\sigma^2} \right) \cos \left( k_\theta^T \delta(x) - \Delta \psi \right)$$

where $|\tilde{I}|^2$ is the constant power spectrum of the input noise images, assuming for the sake of simplicity that locally $I_L \approx I_R = I$, and $k_\theta = (k_\theta \sin \theta, k_\theta \cos \theta)$ is the frequency vector. From Eq. 3, considering $k_\theta^T \delta(x) = k_\theta^T \delta^0$, where $\delta^0$ is the projection of the full disparity along the orthogonal direction with respect to $\theta$, we observe that the complex cell is tuned to a specific stimulus disparity $\delta^0 = |\Delta \psi|_{2\pi}/k_\theta$, depending only on the phase difference between the left and right RFs, $\Delta \psi = \psi_L - \psi_R$ (see Fig 1). Since the phase is constrained to its principal value in the interval $(-\pi, \pi]$, the maximum disparity to which the cells can be selective is $\pm \Delta = \delta^0_{\text{pref}} |\Delta \psi|_{\pm \pi} = \pm \pi/k_\theta$. Grounding on the phase-shift model, we followed the specification of [25] to construct a population of $N_p \times N_o$ disparity detectors with $N_p = 9$ phases and $N_o = 8$ orientations, equally spaced between $-\pi$ and $\pi$ and between $0$ and $\pi$, respectively. The population, in such a way, allows us to obtain an efficient and complete representation of the visual signal with a reduced number of channels. Though the complex cell tuning curve presents a peak of response in the preferred disparity, from Eq. 3 we evince that it is also proportional to the energy of the image $|\tilde{I}|^2$. Using a divisive normalization step [3], [13], we can rescale the activity of a single cell by the energy $E_{bin}$ of the entire population, pooled over all the phases and the orientations:

$$E_{bin}(x) = \frac{1}{\pi} \int_0^\pi \int_{-\pi}^\pi r_c(x; k_\theta, \Delta \psi) d\Delta \psi d\theta$$

$$= \frac{1}{\sigma^8} \left( 1 + \exp \left( -\frac{|\delta(x)|^2}{2\sigma^2} \right) \right) |\tilde{I}|^2.$$

The resulting population response eventually oscillates about an unitary mean. This normalization stage allows to eliminate the dependency on the stimulus luminance, preserving the tuning on the stimulus disparity.

### III. THE LEARNING ALGORITHM

The goal of the learning algorithm is to derive a proper disparity-vergence control signal $v_s$, by a weighted linear summation of the population response in a foveal neighbourhood $\Omega$, so that:

$$v_s = \sum_{x \in \Omega} \sum_{i=1}^{N_p \times N_o} w_i r^i_c(x)$$

In [7] we demonstrated that the network, for its implicit characteristics, can be exploited to drive the progress of learning and to move the system towards an effective behaviour. The internal state of the population can be measured by dispersion of the energy of the activation of the complex cells from the mean, i.e. by the standard deviation (STD) of the population activity. The divisive normalization, by removing the dependance on the stimulus energy, allows for a mean population activity that is almost constant about a unit value. The STD is a function of the modulus of the stimulus disparity. From Eq. 3 we can derive the STD for a
white noise stimulus:

\[
\text{STD}\left(\delta\right) = \sqrt{\frac{1}{\pi} \int_{0}^{\pi} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\bar{r}_e(k, \Delta \psi)\right)^2 d\Delta \psi d\theta - 1} \quad (6)
\]

\[
= \sqrt{\frac{1}{1 + \cosh \left(-|\delta|^2/\sigma^2\right)}}
\]

In order to verify and validate the stability of this measure, we computed the response profiles, and thus the STD, not only for a white noise stimulus (see Fig. 2, black line), but also to more natural and real stimuli, i.e. to pink noise (magenta line), and to a natural image (blue line). Having a maximum for zero disparity and decreasing as a function of the disparity, this signal can be used as an effective reward for the system, both to indicate how much the system is close to the desired vergence (zero disparity) and to evaluate how much effectively and rapidly it guides the action toward fixation. From this perspective, the STD signal reflects the variability of the population activity, thus providing a proper measure of the distance from the goal of the vergence movements, which eventually moves the system toward zero disparity and thus toward the maximum variability.

Differently from [7] where we rely on a particle swarm optimization algorithm for the development of the weights, which is far from being a plausible cortical mechanism, here we adopt a more biologically plausible Hebbian rule to learn the vergence control. Specifically, since the effect of the performed vergence movements can be evaluated at an instant of time subsequent to the one that triggered it, we introduced a temporal asymmetry in the algorithm obtaining a differential Hebbian rule [24]. Analyzing the temporal trend of the population it turns out that the vergence control directly correlates with the variation of the activity of the complex cells. This observation leads to the rule for the update of the weights \(\bar{w}_i\):

\[
w_{i|t} = (1 - \eta)w_{i|t-1} + \eta \text{VS}(r_c_{|t-1}) \Delta r_c^i \quad (7)
\]

where the single synaptic weight \(w_i\) at the instant \(t\) is modified by its value at the previous instant \(w_{i|t-1}\) plus the update term. Such a term is composed of the activity of the postsynaptic neuron (i.e. the vergence control) computed at the previous instant \(\text{VS}(r_c_{|t-1})\), multiplied by the variation of the activity of the presynaptic neuron (complex cell) \(\Delta r_c^i\) elicited by the vergence control. The parameter \(\eta\) models the learning rate. Since the variation of the cell’s activity can be positive or negative, so as the vergence control, such a rule positively reinforces the synapses of presynaptic neurons whose variation correlated to output of the postsynaptic neuron, and negatively reinforces those with anti-correlated variations.

In order to avoid the weights to diverge to infinity, they are divided by a normalization term, as in Oja’s rule [19]:

\[
\bar{w} = \sum_{j=1}^{N_c \times N_s} \left( [(1 - \eta)w_j_{|t-1} + \eta \text{VS}(r_c_{|t-1}) \Delta r_c^j]^{p}\right)^{1/p} \quad (8)
\]

with \(p\) set to 2. Finally, the weights used in Eq. 5 to compute the VS during the learning stage are \(\bar{w}_i = w_i/\bar{w}\).

At the beginning of the learning phase, the weights are randomly initialized with uniform distribution between -1 and 1 with zero mean. Consequently, the initial vergence control is far from producing the correct movement. Nevertheless, the learning rule enhances the influence of the cells with correlated and anticorrelated variation, no matter what is the result of the action taken. Instead of a fixed learning rate, our purpose is to include a dynamic rate that is derived from the activity of the population, the variation of STD, i.e. \(\eta = \eta(t) = \Delta \text{STD}\), that modulates the learning algorithm [14], leading to:

\[
\bar{w}_i|t = [(1 - \eta)w_{i|t-1} + \Delta \text{STD} \text{VS}(r_c_{|t-1}) \Delta r_c^i] / \bar{w} \quad (9)
\]

If from a time step to the next, the STD grows, it implies that the vergence movement has been correct, and the learning rule increases/decreases the weights corresponding to cells with correlated/anticorrelated variation. Otherwise, the vergence movement is incorrect, the sign of the learning rate is opposite, as well as the update of the weights.

The proposed rule, considering a reward term directly in the learning, provides an update of the weights that depends directly on the resulting action in two ways. First, qualitatively because the sign of the variation of the reward \(\Delta \text{STD}\) inverts the contribution to the synapses. Second, quantitatively because the weights are enhanced proportionally to the effectiveness of the action, i.e. on the modulus of \(\Delta \text{STD}\). The resulting algorithm rapidly converges to a vergence control able to move the system toward zero disparity, and to provide stable and correct fixations.

An interesting feature of the computational substrate to be taken into account is that the STD is effective in driving the learning because of the characteristics of the tuning curves. Indeed, if the network were composed of cells with ideal response (Gaussian tuning curves), the same signal would be flat in the region of interest (see Fig. 3, magenta line), providing no information about the state of the system, and thus being useless for the learning.
The learning algorithm is tested in three different conditions. The first experiment is conducted with synthetic stimuli characterized by constant disparity, and addresses the capability of convergence to a proper solution, and the velocity of convergence with the different learning signals. The second experiment analyzes the development of the vergence movement in a realistic situation, i.e. while the eyes are looking at a plane. In the third experiment we test the robustness of the algorithm by using a reduced subset of cells, chosen randomly from the original population.

A. Learning on synthetic stimuli

The algorithm was first tested with synthetic stimuli randomly selected among those used to derive the population tuning curves in Sec. II (RDS, pink noise images and natural images). The binocular stimulus is created from two images, where the left is an equal version of the right one, shifted by an equal amount to the corresponding binocular disparity.

Such a configuration can be considered a facilitated one, because the disparity is directly controllable and constant over the whole image, and it is used to demonstrate the learning capabilities of the network. In order to have a direct comparison with other methods proposed in literature, the algorithm was tested using as reward three different signals (see Fig. 3), i.e. the standard deviation of the population response (STD), the ground truth disparity (GT), and the response of a zero disparity complex cell with \( \theta = 0 \) and \( \Delta \psi = 0 \). For each trial, the system is left moving for a number of time steps that are in principle sufficient to complete a vergence movement.

Fig. 4 shows the evolution of a set of weights using the standard deviation of the population response as the reward signal. The effectiveness of the weights along their development can be assessed from the tuning curves of the resources. Indeed, the resulting vergence control can be derived by a weighted summation of the tuning curves. Since at the beginning of a repetition the weights are initialized randomly, the vergence control derived has not a specific shape, both in term of the 2D profile and for what concerns its horizontal cross section (see Fig. 4, top left). As a consequence, the resulting control, instead of moving the system toward zero disparity, produces wrong vergence trajectories (see Fig. 4, top right). After a limited number of trials, the profile of the control evolves in a more symmetric shape and produces vergence trajectories that drive the system toward zero disparity (see Fig. 4, middle). At the end of the learning process, the system is able to complete effectively and precisely the vergence movement within the given number of time steps (see Fig. 4, bottom).

A qualitatively analysis of the learned set of weights and their capability, pointed out that the obtained profile is characterized by two salient features typical of an effective vergence control: it develops an odd symmetric control with a zero crossing for zero disparity (see Fig. 5, left). At a first glance, the learned set of weights have a characteristic structure that appears after a sufficient learning time. Indeed, so as the control, the weights exhibit a symmetry with respect to zero phase shift and vertical orientation (see Fig. 5, right). Cells characterized by the same orientation provide opposite contribution with respect to the zero phase shift, so as cells with the same phase shift provide opposite contribution with
respect the vertical orientation. It is thus evident how cells with an odd tuning curve ($\Delta \psi \neq 0, \pm \pi$) contribute and compete in realizing the control, whereas cells with even tuning curve ($\Delta \psi = 0, \pm \pi$) provide null contribution, as well as cells tuned to pure vertical disparity.

Each reward signal was tested for different disparity ranges: $[-\Delta/4, \Delta/4]$, $[-\Delta/2, \Delta/2]$, $[-3\Delta/4, 3\Delta/4]$ and $[-\Delta, \Delta]$. To evaluate the mean behaviour of the learning evolution, the experiment was repeated 50 times, and each repetition evolves over 1500 different trials. To compare the velocity and precision of algorithm for different learning signals, we computed the residual vergence error at the end of each trial, i.e. the residual disparity (see Fig. 6). From the evolution, it turns out that the larger is the range of disparities the system has to cope with, the longer it takes to converge to an effective control. In particular, by considering the smallest range ($\pm \Delta/4$), the algorithm is able to develop an effective vergence control within the first 20 trials, providing a good accuracy of the final position (residual error smaller than $\Delta/8$). As expected, whereas the STD and the $\delta_{GT}$ are able to develop an effective control in all the tested ranges, the PH0 signal takes a slightly longer time and it is not able to make the system to converge for a large disparity range ($\pm \Delta$). This happens because PH0 is not monotone and it changes sign at about $\delta = \Delta/2$ (see Fig. 3), providing a correct reward for small ranges of disparity only.

Comparing the STD and $\delta_{GT}$, the velocity of convergence can be considered equal for the two signals, thus validating the proposed unsupervised approach. The system, by exploiting an internal parameter, the STD, is able to learn a correct and effective vergence behaviour, with similar capabilities as if it were exploiting a supervised signal, i.e. the $\delta_{GT}$.

B. Learning on a frontoparallel plane

We consider a virtual environment in which the eyes, characterized by null version and elevation angles, and by a vergence angle $\alpha$, look at a frontoparallel plane with a pink noise texture (Fig. 7). The plane is at a depth $Z$ with respect to the cyclopean position, and perpendicular to the binocular line of sight. The interocular distance is $b = 70mm$, the focal length is $f_0 = 17mmm$, and the stimulus is projected onto the retinal plane, with a size of $6mm$, thus considering a field of view of about 20 degrees. The vergence control is computed over a spatial neighborhood defined by a Gaussian profile centered in fovea and with standard deviation of $1.5^\circ$.

![Fig. 7. Simulated experimental setup for learning the vergence on a frontoparallel plane.](image)

The geometry of the binocular vision system is characterized by common tilt for the left and right cameras ($\alpha_L$, $\alpha_R = \alpha_L$), and independent pan angles ($\beta_L$, $\beta_R$), as in the Helmholtz reference frame [10]. This configuration yields a simplified parameterizations of the visual direction in terms of version ($\upsilon$) and vergence ($\chi$) angles [6]:

$$\begin{cases}
\upsilon = 1/2(\beta_L + \beta_R) \\
\chi = \beta_L - \beta_R
\end{cases} \quad \text{or} \quad \begin{cases}
\beta_L = \upsilon - \chi/2 \\
\beta_R = \upsilon + \chi/2.
\end{cases}$$

Hence, the vergence control needed to move the fixation point, while keeping constant the gaze direction, is a quantity $\Delta \chi$ to be applied symmetrically on both the eyes: $\Delta \chi = \Delta \beta_R = -\Delta \beta_L = -\arctan(\frac{\Delta \upsilon}{2Z})$.

Whereas with parallel optical axes, the horizontal component of disparity is a function of the depth of the stimulus, and the vertical one is zero across the whole image, considering a realistic system with vergent geometry, the arising pattern is more complex [8]. In particular, adopting the Helmholtz geometry, the optical axes are always intersecting, and the vertical disparity is zero in the fovea only. Indeed, the vertical disparity grows with the retinal eccentricity, and it is enhanced by three factors: (1) - by a gaze direction different from the straight-ahead, (2) - if the observed surface is not perpendicular to the binocular line of sight, (3) - for large vergence angles. Thus, seeking to learn a vergence control that would be able to cope with a changeable and unknown environment, the vertical disparity is not negligible, and it is mandatory to take it into account. In order to verify the
effectiveness of the learning algorithm in a more realistic situation, we moved the gaze direction randomly from \(-20^\circ\) to \(20^\circ\) for the azimuth, and from \(-20^\circ\) and \(20^\circ\) for the elevation. Moreover, the orientation of the plane in space is not perpendicular to the binocular line of sight, but it is tilted and slanted of a random angle between \(-10^\circ\) to \(10^\circ\).

At the first time step, the plane and the fixation point are at the same \(Z\), then the plane is moved to a new depth, and the vergence angle starts to moving. Working in a visually-closed loop, the control is updated at each time step and, particularly for small disparities, it has a value proportional to the binocular disparity in the fovea. As a consequence, when the fixation point reaches the depth of the stimulus, the vergence control intrinsically turns to zero and the fixation point remains steady at the depth of the plane.

To demonstrate the effectiveness of the control learned from a real plane, we show three different tests with a frontoparallel plane whose position in depth varies in time as a step, as a ramp and as a sinusoid (see Fig. 8). These tests show how the system developed the two salient features of the control. In fact, it is capable of producing fast changes of the fixation point for abrupt changes of the stimulus depth (Fig. 8, top) and smooth movements to follow the depth of the stimulus (Fig. 8, center and bottom), evidencing the odd symmetry of the control. Likewise, it is able to produce as stable fixations on a steady stimulus (Fig. 8, top) with the fixation point laying correctly at the depth of the stimulus, testifying the existence of the zero crossing for zero stimulus disparity. In the same way, the frequency of the sinusoid that determines the temporal variation of the depth of the plane, was varied between 7 and 38 time steps (not shown). Again the simulation results are qualitatively similar to the experimental data [11]. The learned vergence control (blue line), ensures both the ability to track a moving stimulus in depth (dotted line), and to provide robust and stable fixations on steady stimuli.

C. Learning with a limited number of resources

Distributing the resources over the feature space may result in an oversimplification of the algorithm, since the system is pushed to develop a symmetric control through a symmetric exploitation of the resources. Indeed, it is unlikely to suppose that in a real neural architecture the resources, although covering the whole feature space, are so regularly distributed over it (see for example [22]). In order to further validate the proposed learning algorithm, we repeated the tests removing from the population a subset of cells (see Fig. 9, right), randomly chosen among the whole set so to prevent the system to rely on symmetric resources. Nevertheless, the developed control provides the two required salient features (see Fig. 9, left), and show a capability that is qualitatively equivalent to that obtained by the whole population.

Analyzing the velocity of convergence of the system when exploiting a subset of the resources (see Fig. 6), it is evident how it takes a longer time to develop an effective control, but the final behaviour is equal to that obtained by the whole set of resources. This consolidates the robustness of the learning algorithm and the effectiveness of the reward signal.

In conclusion, by exploiting a precise and extensive knowledge either of the environment (ground truth disparity) or of the resources (tuning curves) allows us to obtain an effective vergence control. Nevertheless, such techniques are grounded on a kind of knowledge that prevents the system to autonomously learn its behaviour. On the other side, the whole activity of the network, designed by mimicking the neural parameters, is an intrinsic parameter of the architecture, which is able to drive the learning of the vergence behaviour and to obtain equivalent performances.

V. CONCLUSION

In this paper, we proposed a neural architecture that is able to exploit its internal state to drive the development a proper vergence behaviour, thus without any external supervision. The proposed approach, grounds on basic cortical mechanisms both for the representation of the visual signal and for the learning strategy.

From an algorithmic point of view, the distributed coding of the visual signal, allows an effective evaluation of the state of the system, and, consequently, of the action taken to reach another state. Relying on the efficacy of the divisive normalization mechanism, the evaluation signal is robust and stable, and invariant with image features other than the disparity. The obtained population response is modulated by the retinal disparity only, both with white noise stimuli and with more natural ones like pink noise and natural images.
More precisely, the estimation of the relationship between the system and the environment (i.e. the system’s state) is obtained on a visual basis. Indeed, it is directly the internal representation of the visual signal that drives the learning of the motor control, causing a behaviour that naturally emerges from the interaction between the system and the environment. The exploration of the environment performed by the system, also corresponds to an exploration of the capabilities of the system itself, and to an autonomous tuning exercise for a better interaction with the environment.

From a methodological point of view, the proposed architecture points out the advantages and the flexibility of distributed cortical-like architectures against solutions based on a conventional systemic coupling of sensing and motor components. In fact, each cell of the population cooperates for the generation of the motor control, attesting how a complete representation of the visual signal is instrumental to gather an effective motor control. The cortical-like architecture is thus an effective substrate, able to derive correctly the vergence movements, and to gather the reward signal for the learning phase. The proposed learning rule, extending the standard Hebbian learning to take into account (1) a temporal asymmetry (and thus the dynamic of the system), and (2) a reward signal that directly affects the learning rate, implement a direct and mutual influence between the control and the population activity that generated it, allowing to develop an effective vergence behaviour.

More generally, the proposed architecture allows us to avoid a hard sequentialization of sensory and motor processes, that is certainly desirable for the development of cognitive abilities, e.g. when a system must learn binocular eye coordination, handling the inaccuracies of the motor system, and calibrate the active measurements of the space around it.

REFERENCES


